EVALUATION OF SPATIAL PRESENTATION IN SONIFICATION FOR IDENTIFYING CONCURRENT AUDIO STREAMS

*Hong Jun Song, *Kirsty Beilharz, †Densil Cabrera

*Key Centre of Design Computing and Cognition
†Acoustics Research Laboratory
Faculty of Architecture, Design and Planning
University of Sydney
Sydney, Australia
hson5790@arch.usyd.edu.au
kirsty@arch.usyd.edu.au
densil@arch.usyd.edu.au

ABSTRACT

The ultimate goal of sonification is to transfer information effectively to listeners. While there is a large amount of multidisciplinary investigation in the field of psychoacoustic, psychology, cognition and human computer interaction, sonification design still lacks empirical evidence on which to base design decisions [1]. This paper presents an empirical investigation of spatialization, which can provide one or more dimensions for auditory display. It focuses, in particular, on evaluating spatial presentation in sonification so as to enhance pattern identification when two audio streams are played simultaneously. Hence it aims to develop design decisions that benefit from effective information representation. The sounds were created for binaural reproduction using non-individual head-related transfer functions. The results reported are based on the listeners' performance within two display modes: (i) two co-located streams and (ii) two streams spatially separated at static locations. It concludes with ideas for future improvements and developments for this type of sonification.

[Keywords: Spatialization, HRTF, Spatial separation, stream segregation]

1. INTRODUCTION

Sonification delivers information through non-speech sound [2]. Due to its intuitive connection with symbolic meanings, sound has been used extensively as a means for Human Computer Interaction (HCI), where it can provide rapid comprehension of the processing status of, alarms or warnings related to a system. With the rapid growth of digitalization in our society, auditory display meets the needs of extending computational methods for processing data, where sound is able to relieve visual overload and provides us with important feedback on our actions. Taking account of ubiquitous and profitable listening to sound, the industry potential of sonification includes reinforcing visualization in bi-modal display situations and providing a non-visual alternative for embedding information efficiently. For this reason, binaural display and non-individual HRTFs are used (for isolation and affordability) that mimics the workplace environment where we envisage this sonification occuring. We recognize that the accuracy of head-tracking, individualized HRTFs or loudspeaker reproduction is greater but these laboratory conditions do not simulate workplaces, such as stock trading data analysis context.

In information sonification, various types of information need to be displayed clearly and unambiguously. Complexity arises with multiple audio streams since they reach both of our ears as a mixture. The streams need to be distinguishable from each other for people to have an adequate understanding of the sonification. In a multi-stream sonification, it is crucial to produce an effective mapping scheme to facilitate the listener's ability to follow individual strands of the message and to understand its overall meaning. The cognitive process of separating individual meanings from the mixture is known as auditory stream segregation. It is related to people's ability to interpret complex auditory scenes according to sound properties such as pitch, tempo and location. Many findings in this field are incorporated into the conceptual framework of auditory scene analysis (ASA) [3].

2. PRESENTING TWO CONCURRENT AUDIO STREAMS WITH SPATIAL SEPARATION

Concurrent presentation increases the information presentation dimensions and allows a parallel processing [4, 5, 6]. However, there are potential difficulties of monitoring simultaneous sound streams. Depending on the number of auditory streams to be identified, the tasks of recognition include divided or selective attention [1, 7, 8]. A divided attention task involves people tracking the changes of two or more stimuli at the same time; in a selective task, only target parameters need to be extracted within the presence of the other competing sources. Interference and masking arise when both of them compete for attention. One example is the phenomenon of binaural interference [9], where binaural judgment of high-frequency signal can be disrupted by simultaneous low frequency signal. Also the listener may not be able to separate one audio stream from the other if they have similar semantic structure (i.e. harmonic, timbre or tempo). In order to activate appropriate auditory stream segregation, differentiation of concurrent audio streams is usually achieved by having separated frequency bands, isolated direction, distinctive timbre or different speed [1, 10].

Spatial separation has a better "force and semantic structure" than pitch to reduce "problems of peripheral sensory masking" [11] and maintain attentions to sound sources. Directional cues, acting as a spotlight, enhance the processing of sounds and speed up discrimination responses by providing an essential interpretive context that gives meaning to sound. On the basis of experience, directional cues aid differentiation of subjective mental representation of audio streams [12]. The benefits are from two features of spatialization. Firstly, spatialization is relatively independent and bears no common forms with other parameters. Thus spatial representation makes reasoning about audio streams "easy". The fundamental frequency and location of a tone can be changed with negligible interaction, whereas if fundamental frequency and sound pressure level are changed independently, ambiguity may occur due to the interaction between these dimensions (heard as pitch and loudness respectively). Secondly, interaural time/level differences reduce masking and interference between streams [13, 14] and improve the signal-to-masker ratio and the audibility of the target that allow listeners to direct their attention to the target.

While intuition may indicate that spatialization enhances auditory stream separation for sonification [15], there is little empirical information in research literature supporting and investigating the effects of non-speech spatialization. Additionally, "binaural interference is a byproduct of grouping processes" which may lead to the location of simultaneous stimuli cannot be perceived accurately [9]. The diminished effect of spatial separation would degrade the efficacy of stream segregation. On the other hand, controversy arises, for instance, that "knowing where the sound is coming from seems to have little help to detection" [11]. When spatial tasks were involved, a decrease of efficiency has been found within visual research [16, 17, 18] (which is usually analogous to auditory display). As shown with the augment of the spatial items in the search array, reaction time to the visual tasks increased and accuracy were reduced. When two sound sources are oriented from different directions, the listener needs to keep tracking and integrating auditory messages from two spatially separated locations. Spatial rehearsal allows tracking the orientation locations in working memory and spatial reasoning is augmented other than masking and interference of two audio streams. Divided attention, involving both competing auditory messages and imposing additional cognitive demands, probably decreases performance [19]. The other argument is that spatial separation has less support for non-categorical information (e.g. words) because the temporal information about pairs of stimuli is destroyed due to the spatial cognitive ability [20]. It seems the positive influence of spatial separation is only restricted to categorical streams for sonification. These arguments caused a hesitation of using separated spatial location for competing audio streams in sonification. This study is motivated by the need to systematically explore and demonstrate the ways in which binaural spatialization influences auditory stream segregation.

3. PREVIOUS STUDIES ON SPATIAL SEPARATION

The influence of spatial separation has been extensively investigated for speech. The cocktail party effect is well-known in the field of speech recognition, referring to the ability of listeners to separate a single talker from competing talkers and background noises, and to concentrate on specific conversation [21]. Research in speech field consistently shows the benefits of spatial separation for splitting competing verbal messages. Speith [22] investigated listeners' responses to one of the two simultaneous speech streams when presenting with loudspeakers. It was found that increased horizontal separations always improve accuracy and multiple channels (horizontally separated loudspeakers) are more effective than one loudspeaker (midline). A related study was conducted by Webster and Thompson [19], showing that when multiple loudspeakers are applied the ability to respond to the streams of sequential messages can be enhanced. Following early work [22, 23, 24], numerous efforts have been dedicated to exploring the effect of spatialization in speech. Shinn-Cunningham et al [25] explored the effect of spatial separation on recognizing the content of the verbal sentences. The results illustrated that overall performance was enhanced by arranging speech signal from different locations although the effect in selected tasks was more significant than in divided tasks. Also, fixed locations were better recognized than randomly altering the position from trial to trial. Other study by Best [14, 26] also demonstrated the advantage of perceived spatial separation (up-down and left-right) between the concurrent pairs of spoken word stimuli.

A study of earcon identification has shown that spatially located concurrent earcons were more easily identified than those having unique location [27]. Participants in the experiment were found to be able to identify significantly more numbers of concurrent earcons in spatial distinct locations, as well as increase the identification of earcon register. The result confirmed that "sufficient separation in space" (more than 60° separation) can improve the effectiveness of spatial separation for concurrent earcons (especially categorical data).

4. AIMS

This paper looks at the possibility of spatial separation to assist stream segregation during deciphering sonification data. Our concern is particularly to develop understanding of the potential application of spatialization in sonification using commercially available binaural reproduction. Non-individualized headphone reproduction, without head tracking, provides a simple and practical spatial audio system which can easily be applied for workplace sonification applications. The current study quantifies the benefits of spatial separation for concurrent audio streams in sonification. The influence of spatial separation is explored by measuring the performance in co-located and separated situations. It shows that, in auditory graph task, spatial separation can enhance pattern identification of two concurrent signals. The results that will be presented in this paper point to the enhancing comprehension of sonification by using spatialization in stream segregation activities. Future experiments aim to improve and extend the findings of timbre and spatial distinction in an auditory display of contextual data sets.

5. METHOD

The hypothesis of this study is that the usage of spatial separation is able to improve the performance of divided tasks. The task is to listen to concurrent audio streams of auditory graph (in terms of an equivalent visual graph) and recognize the content both. The performance is evaluated by comparing the numbers of correctly recognized auditory contours within two spatial layouts: a single horizontal position located in midline or two spatially separated horizontal positions. In both cases, the positions were fixed.

5.1. Participants

Altogether thirty-four volunteers were recruited for this study. Most of them were enrolled students in the "Sound Design and Sonification" class at University of Sydney but they had not yet been exposed to discussion on spatialization at the time of the test. Data

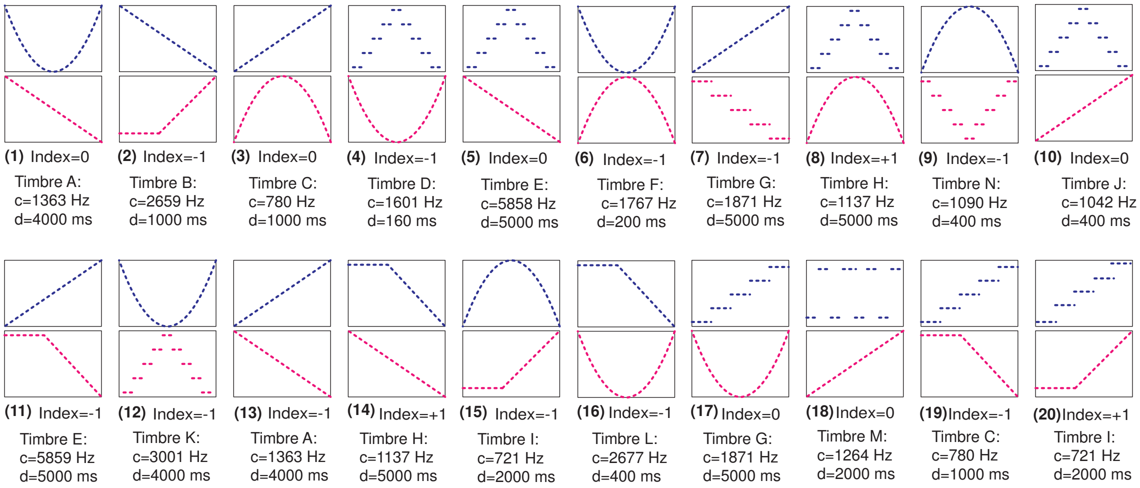

Figure 1: *Twenty pairs of graphs for sonification numbered from (1) to (20). Values of each graph were mapped onto pitches and each graph represents one audio stream. A set of timbre (A - N) were employed for each pair of graphs, where 'c' is the spectral centroid in Hz and 'd' is the duration of attack phase in ms (including attack, sustain and decay). The index shows the correlation between two auditory graphs where '-1': negatively correlated,'+1': positively correlated and '0': no correlation. The index was generated according to the Pearson correlation coefficients and significant values (p-value).*

from two participants has been excluded: one due to an accidental equipment problem; and another due to excessive deviation from the group mean (this participant reported not reading the instructions carefully). The selected participants ranged in age from 19 to 41 years with a mean age of 23 years and there were 15 females and 17 males. All volunteers provided their informed consent according to the Ethics Committee of the university.

5.2. Stimuli

The auditory stimuli were created in Max/MSP [28]. Max/MSP is a real time graphical programming environment. The parameters such as playing tempo can be easily altered by editing the graphic icons. Its combination with visual display is useful for training or demonstrating audio outcomes for participants after the experiment. The SPAT library for Max/MSP [29] supports the HRTF function for binaural synthesis. It is a spatialization library in which artificial reverberation, localization of sound source and spatial content of the room effect are integrated in a single processor patcher. It allows flexible and precise control of these effects [29], in contrast to simple level-based panning.

The experiment used headphones for binaural reproduction with generic HRTFs. No head-tracking was involved. While this simple binaural presentation technology has the two problems of cone-of-confusion errors (such as front-back confusion) and head-locking of the sound-field (the sound-field moves with the listener's head), the spatial separation used for the stimuli was lateral (and so relies predominantly on binaural difference cues, which are conveyed effectively using this technology). Spatial rendering was simply in terms of image direction, without any attempt to vary distance or other aspects of auditory space.

5.2.1. Stimulus generation

Twenty pairs of graphs were selected for sonification, and the concurrent audio streams are represented in Figure 1 using a pair of vertically stacked graphs. Pairs of graphs are combinations of simple curve lines, straight linear lines or steps. The values of graphs were mapped onto MIDI notes ranging from 42 to 91, in which the x-axis of the graph was mapped onto time and y-axis was mapped onto midi note value. The speed was set to 10 points/second and the duration of each stimulus was about 10 seconds. The midi notes in each stimulus were calculated according to the equation (1) proposed by Brown et al. [30]:

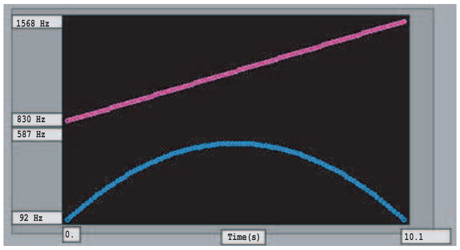

Figure 2: *The x-axis of the graph was mapped onto time and y-axis was mapped to MIDI note value. The pitch rages are 92 to 587 Hz and 830 to 1568 Hz, respectively.*

$$Pitch = note_{min} + \frac{value - val_{min}}{val_{max} - val_{min}}(note_{max} - note_{min}) \tag{1}$$

Then the midi note was converted to frequency in a logarithmic scale in order to be used for oscillator. Pitch register was applied to separate the audio streams. For the two concurrent pitch contours, one was located in pitch range from 92 to 587 Hz and another was from 830 to 1567 Hz (Figure 2).

Each pair was played twice in the two different display modes (Figure 3). In the single sound source condition, both of the concurrent audio streams are from the same sound source, which is 0° azimuth (straight ahead) in the horizontal plane (elevation 0°); or they were spatially separated. Best et. al found that interfer-

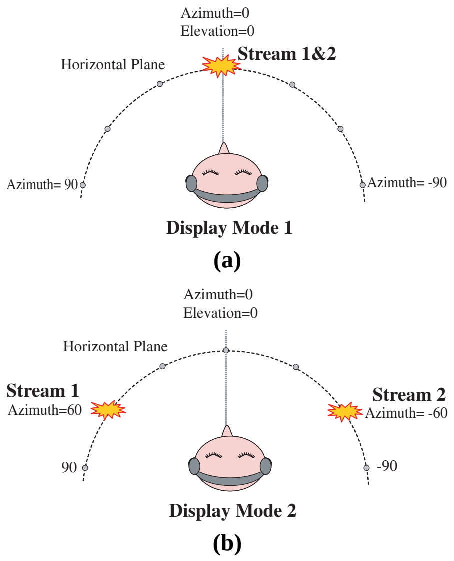

Figure 3: *Spatial configuration. Mode 1: both streams are from the middle line (azimuth= 0°and elevation=0°); Mode 2: they are spatially separated, 60 degrees away from the middle line symmetrically.*

ence still occurred up to 60° separation of verbal messages [26]. According to the previous findings [10, 31, 32] for concurrent minimum audible angle (CMAA), our case used two concurrent sound sources 60 degrees away from the midline symmetrically in the horizontal plane, originating from 60° (to the left of the midline) and -60° (to the right of the midline).

5.2.2. Frequency/amplitude modulation

Pure tones are often not well-suited for auditory graphing because they are difficult to localize, and they can easily fuse into a single auditory image when their frequencies have a simple ratio. The difficulty in localization comes from the facts that rich spectral content is required to make use of pinna-related spectral cues, and that binaural difference cues are ambiguous for a significant range of pure tone frequencies [33]. Fusion can occur because the auditory system tends to interpret harmonic spectra as single pitches (a phenomenon referred to as “virtual pitch“ by Terhardt [34]. Therefore, complex tone stimuli were used in this experiment, synthesised through frequency and amplitude modulation, and using an attack-decay-sustain-release envelope function.

5.2.3. Binaural synthesis

Sounds were presented at particular azimuth angles using generic head-related transfer functions (HRTFs), that approximate pinna-related spectral transformations and interaural time and level differences. The HRTFs provide listeners with distinct and localizable sound source. In this study, we used generic HRTF codes that have been measured on the KEMAR dummy-head, collected by MIT’s Media Laboratory [35]. The 710 points were sampled from -40° to 90° at the elevation in an anechoic chamber at sampling rate of 44.1 kHz. The left and the right channels in this experiment were convolved with HRTFs corresponding to the current direction of the virtual sound source with respect to the listener position. Interaural differences were added to the output channels according to the positional information of the sources. The cues of virtual sounds such as position, direction, distance, orientation and room effect were controlled with SPAT.

5.3. Tasks and procedure

The tasks were to listen to the graph sonification containing a pair of simultaneous audio streams and match the contour of both streams to the visual representations of Figure 1. Those binaural stimuli were presented with a CD player using six pairs of dynamic open-air Sennheiser HD 433 headphones. Those twenty pairs of graphs were displayed as co-located sound sources and as spatially separated sound sources so altogether there were 40 concurrent audio displays. They were arranged in three random sequences to lessen the effect of sequence. Participants were divided into three groups of almost equal size, corresponding to one of the three random sound sequences. The purpose of the experiment was not mentioned in order to avoid listeners being distracted by seeking spatial cues. Instead, before starting the experiment, participants were provided with a one-page instruction sheet that included basic information about the experiment such as tasks and the duration. All participants received equal information from the paper instruction. The first five listening examples were regarded as training and not used in comparison.

For each trial, the auditory stimuli consisted of a pair of contours (10 s) and a repetition. After the repetition, listeners were required to circle one correct answer from 12 options in 7 seconds. Each trial started with a male voice announcing the question number and there was a beep prompt tone before the replay.

After the experiment, participants were required to fill in a questionnaire, in which they were asked to self-evaluate their musical background (MBG), indicating whether they noticed the different spatial display and providing basic personal information.

6. RESULTS

With regard to the aim of this study, the overall performance showed that spatial separation does improve listeners’ ability to attend to two competing pitch contours when using binaural representation. The effect of spatializsation was quantified by comparing the difference of performance when the pair of audio streams were spatially separated or co-located. As the purpose of this study was not mentioned before the experiment, most listeners stated in the post-experiment questionnaire that they were not aware of the spatial separation and were not conscious of the change of the spatial cues. In the survey, a few participants stated that they perceived the difference of spatial display mode and intuitively they thought in spatially separated display the concurrent contours were more easily recognized than for co-located.

The thirty-two subjects are grouped into 4 categories: better, same, worse or no wrong answer according to their performance (shown in Figure 4). If the wrong answers in a single sound source (midline) exceed those for spatially separated sound sources (equally spaced along the azimuth), the sample belongs to category of “better”, which means spatial separation enables better discrimination between two concurrent audio streams. If the wrong answers in the two display modes are the same in number, the sample is classified as “same”, which means the discrimination is not evident. If the wrong answers in the single mode are fewer than in the separated display, the sample is classified into “worse”,

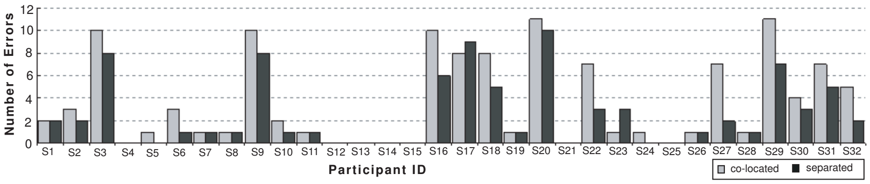

Figure 4: *Performance of 32 subjects. A relatively significant differences show in subjects such as S3, S6, S9 etc.; some participants made no wrong answers at all (e.g. S4 and S12); only S17 and S23 made more wrong answers in the separated display than in the co-located display.*

which means spatial separation did not enhance deciphering. People who did not make any wrong answers (total wrong answers = 0) are in group of "No wrong answer". Group behavioural data are summarized in Table 1.

Table 1: *The performance when corresponding to two concurrent audio streams.*

	Frequency	**Percent**
"better"[1]	16	50.0
"same"[2]	7	21.9
"worse"[3]	2	6.3
"no wrong answer"[4]	7	21.9

[1] Errors (mode 1: single sound source) $>$ Errors (mode 2: spatially

[2] separated sound source) Errors (mode 1: single sound source) $=$ Errors (mode 2: spatially separated sound source)

[3] Errors (mode 1: single sound source) $<$ Errors (mode 2: spatially separated sound source)

[4] Errors (mode 1) $=$ Errors (mode 2) $= 0$

The two groups of wrong answers (each group for one display mode) are dependant and each set of paired wrong answers is from the same sample/subject. A paired t-test compared each set of pairs and analysed a list of difference between two groups. Therefore, their performance [$t(31) = 3.968$, $p = 0.005 < 0.01$] illustrated a significant difference between the two display modes. One sample t-test [$t(31) = 3.968$, $p < 0.01$] showed that the difference of population between the "better" group and other groups is significant, which means performance is better when sounds were emitted from two fixed separated locations than when they were from a single location. The result confirms the utility of spatial separation for concurrent audio stream in divided tasks and it indicates that spatial separation can be used in mapping to distinguish simultaneous concurrent data streams in information sonification when monitoring competing information streams is required.

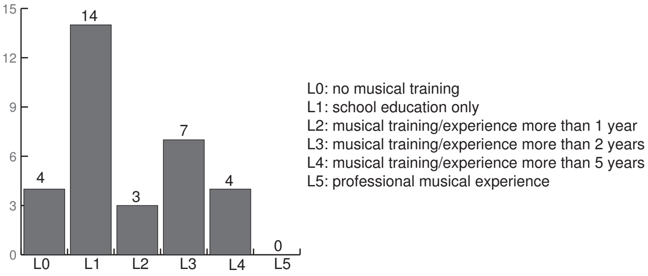

Figure 5: *Musical background (MBG). Six levels were rated from beginner to expert. Numbers of people at each level are marked on corresponding bar. No participants had professional musical training.*

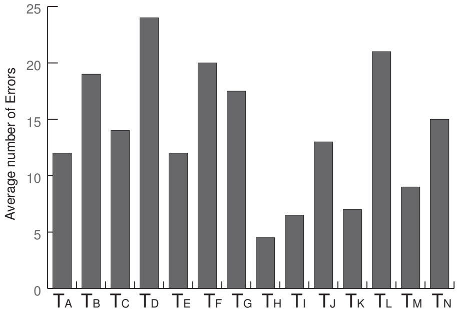

Figure 6: *Average errors of 14 timbres in both spatial display modes. Timbre D, F and L produce the highest number of errors, while timbre H, I and K have the fewest errors.*

7. DISCUSSION

The findings of musical background, configuration of tone colour combined with duration and graph correlation, and gender are discussed in the following subsections. The investigation is useful on which to base design decisions in future work, especially the influence of musical background and tone colour. It provides evidence to determine effective strategies in training session and to compose timbre in mapping scheme. It is uncertain what listening strategy participants were using in this experiment (divided or selective attention) because the stimuli were repeated once. Our intent was to reduce the complexity of the tasks but this might make it possible that listener might be aware of the layout of two audio streams and then concentrate on one stream at the first time and on the other at the second time, especially after several questions.

7.1. Does musical background influence listener's performance?

Musical experience was considered as one of the human factors (such as age and gender) related to cognitive ability in many previous studies. In Neuahoff et al's case [36], musical experts and musical novices responded differently when the tasks was pitch magnitude estimation. Prior musical knowledge and expertise contributed to reaction time and the accuracy of the interpretation. The differences between expert and non-expert occurred in such area like memory, selective attention and categorization and they suggested that "if frequency change is to be used as a dimension to represent a variable in a display, then the changes in frequency employed should be sufficiently large in order to minimize errors in judging the direction of change." Such findings have not been consistent being not sure the role of other individual cognitive difference factors in auditory interpretation. By replicating the Neuhoff et al's approach of assessing musical experience, for the tasks of auditory graph interpretation, Walker and Mauney indicated that musical background is not a significant contributor [37]. We were interested to discover if there was any correlation between musical knowledge and interpretation of auditory graphs with the augment of binaural cues, as we speculated that people with high level of musical training would make fewer errors. Participants pointed out that they became familiar with the display of the stimuli and felt comfortable after first a few trials, so the first five trials were excluded from analysis and regarded as training. In the post-questionnaire, their musical background (MBG) was rated by themselves at six levels from beginner to expert (Figure 5). Most participants are at level 1, in which they have school education only. Four people stated that they have more than two years of music training. None had professional musical experience.

One-way ANOVA has found no significance among six levels [$F(4,27)=0.352>0.05$]. This result indicates that musical background does not significantly influence their performance. Then group 0, 1, and 2 were re-grouped as low-level musical experience, and 3 and 4 as relative high-level group. The comparison of the mean of these two groups [$t(30)=1.447$, $p(0.158)>0.05$] also shows that the level of musical experience does not influence the pattern recognition. These results are consistent with the findings of Walker and Mauney [37].

7.2. Do tone colour and graph correlation influence the performance?

In the experiment, timbre was used to characterize signals, combined with duration (or articulation). Timbre refers to the auditory quality of sound and is defined as "an attribute of auditory sensation, in terms of which a listener can judge that two sounds similarly presented and having the same loudness and pitch are dissimilar" [32]. Many studies have confirmed the important role of timbre for stream segregation, such as ASA [3]. It was considered important that the task remains a purely spatial one in order to faithfully measure the effect of spatial separation. Therefore, to avoid confound of binaural and timbral effects the simultaneous stimuli were using the same timbre and concurrent stimuli only differed in their virtual sound source locations so binaural effects of separating concurrent streams could be isolated from timbre effects.

Although tone colour in this study is not the factor to differentiate the two concurrent audio streams, it contributes to sound quality. Tone quality potentially affects masking of concurrent signals and some timbres were perceived as more distinctive than others in the questionnaire responses. This paper considers the role timbre plays in sound perception that may be useful for future improvement, although timbre was not an independent parameter and its impact was confounded with the pitch contours.

Fourteen timbres were employed in this experiment by altering modulation index, the ratio of frequency of the carrier to the frequency of the modulator and oscillator envelope. The comparison of total errors occurred in both of the two display modes (Figure 6), shows that timbre D, F and L were perceived worse than others and timbre H, I and K were better. Observation of timbre features and duration is based on the analysis of spectral centroid. There are other techniques such as tristimulus [38] (which describes timbre equivalent to colour attributes in vision) and spectral irregularity/smoothness [39, 40] (in which the average of current, next and previous amplitude is compared with current amplitude). The spectral centroid corresponds to the brightness/sharpness of a sound, which is perceived brighter with more high frequency components. According to the ranking of average errors for each timbre in Figure 6, Figure 7 demonstrates the extracted features of the three best and worst performing timbres. The results are the time dependant distribution of the centroid of the signal. It is found that listeners usually performed worse when attack was shorter and the decay was faster, such as timbre D, F and L. Except timbre H, the correlation index of other two timbres (I and K) were "-1". The ranking indicates that with the timbres of relatively longer attack period, the positively correlated pair of auditory graphs had fewer errors than negatively correlated pairs.

7.3. Does gender influence the performance?

There are almost equal numbers of male and female samples (15 female and 17 male). The factors of gender and age are often examined in relation to their impact on performance. Both male and female groups have samples which made no wrong answers (6 and 2 people). The result of ANOVA suggested that there was no difference in performance between female and male listeners [$F(1,30)=0.090>0.05$].The Spearman correlation test showed that it is likely that people, who cannot do well in spatially separated display, also cannot perform well in co-located display.

8. FUTURE WORK

The results indicate that spatial separation can be a valuable method to separate pairs of audio streams in a graph sonification. Our future work includes the improvement of aesthetic and technical presentation in detail and an application of sonification in a "real life" context.

Binaural headphone sound design will be optimized through three aspects. (1) Equalization of headphones could lead to a better spatial impression although it is a subtle key for optimizing binaural reproduction [41]. Prior to headphone presentation, the head-related transfer function can be equalized for Sennheiser HD 433 according to its manufacturer's designed impulse response. (2) Timbre influences the quality of sounds. Listenable (over time) and effective sound is an important consideration in designing sonification for uses in practical situations. As shown in the results analysis, inaccurately identifying the contours often occurred in the sound sequence with a harsh tone colour whether the sound sources were spatially separated or co-located. Our following studies will try to utilize efficient and distinctive yet "palatable" timbres to make the auditory display attractive and maximally comprehensible. (3) In the task of contour identification, lack of training does not cause problems. But in the future if complex tasks, i.e. monitoring multi-stream large scientific data sets, are involved, there is an increasing demand for deliberate training strategies for users. We assume that training on specific tasks will decrease the difficulties of the monitoring or exploratory tasks and improve the performance, according to the previous studies on human capability of auditory learning and adaptation [42, 43, 44]. Based on the role of training and lack of musical experience of participants, the future experiment will apportion greater time to training.

Future work will apply our findings in a sonification context. The potential application includes finding an effective way to explore or monitor data by using sound. Combining with other com-

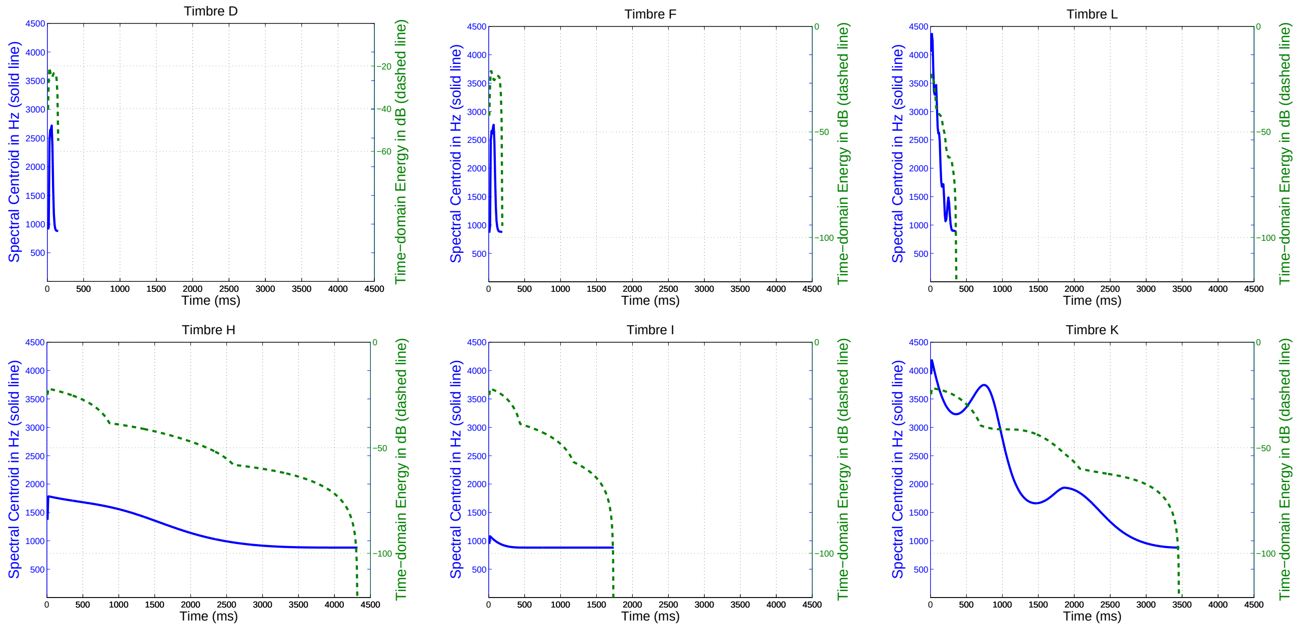

Figure 7: *Feature extraction of an 880Hz tone with the most effective and least effective timbres, where FFT size=1024 (Blackman windowing) and hop-size=512. The attack periods (attack, sustain, decay and release) of Timbre D, F and L were shorter than H, I and K.*

munication modes such as visual display, bimodal display will be considered but auditory display will be our priority focus. The visualization schemes would only supplement and augment sonification rather than distracting listeners. The representation system of complex data sets (large and highly dimensional) will try to satisfy aesthetic and functional requirements. Effective sonification will be relatively intuitive to interpret. The data will be digitalized information or activities, such as motion tracking and environmental data for our sensate lab “Curious Spaces” project. These kinds of numerical data normally have multiple information dimensions. Two prominent sets, which contain important patterns, will be selected for sonification. The sonification strategy used will be quite simple. The values of the information will be scaled to match human sensory characteristics and then mapped onto auditory dimensions. The mapping scheme will be designed depending on the tasks and data variables selected. The sound sequences of two data dimensions are virtually located in two spatially separated sound sources, presented to listeners via headphones, and virtual locations will be simulated through binaural reproduction with generic HRTF. The purpose of the proposed sonification system is to continue to qualify the benefit of spatialization in a sonification context. To encourage participants to follow the sound, the idea of including slight interaction will be considered, i.e. participants can trigger a human computer interaction (HCI) when they find any primary feature. The performance will be measured by examining the primary features revealed by participants before and after spatialization is augmented. Furthermore, the creative aspect of sonification can facilitate an interactive experience which is able to provide educational benefits.

9. REFERENCES

[1] J. Anderson, “Creating an empirical framework for sonification design,” in *11th International Conference on Auditory Display (ICAD’05)*, Limerick, Ireland, 2005, pp. 393–397.

[2] G. Kramer, “An introduction to auditory display,” in *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed. 1994, vol. XVIII, pp. 1–77, Addison-Wesley.

[3] A. S. Bregman, *Auditory scene analysis*, The MIT Press, Cambridge, Massachusetts, London, England, 1990.

[4] C. Hayward, “Listening to the earth sing,” in *Auditory Display: Sonification, Audification and Auditory Interfaces*, Gregory Kramer, Ed. 1994, vol. XVIII, pp. 369 – 404, Addison-Wesley.

[5] R. Minghim and A. R. Forrest, “An illustrated analysis of sonification for scientific visualization,” in *Visualization 95*. 1995, pp. 110–117, IEEE.

[6] T. W. Nattkemper, T. Hermann, W. Schubert, and H. Ritter, “Look & listen: Sonification and visualization of multiparameter micrographs,” in *25th Annual International Conference of the IEEE Engineering in Medicine and Biology Science (EMBC2003)*, Cancun, Mexico, 2003.

[7] S. Haykin and Z. Chen, “The cocktail party problem,” *Neural Computation*, vol. 17, no. 9, pp. 1875 – 1902, 2005.

[8] J. Anderson, “Sonification design for complex work domains: Streams, mappings and attention,” 2005.

[9] V. Best, F. J. Gallun, S. Carlile, and B. G. Shinn-Cunningham, “Binaural interference and auditory grouping,” *Journal of the Acoustical Society of America*, vol. 121, no. 2, 2006.

[10] K. M. Franklin and J. C. Roberts, “Pie chart sonification,” in *Information Visualization*, Ebad Banissi, Ed. 2003, pp. 4–9, IEEE Computer Society.

[11] H. Hawkins and J. Presson, “Auditory information processing,” in *Handbook of perception and human performance*, KR Bott, L Kaufman, and JP Thomas, Eds., vol. II of *Cognitive processes and performance*, pp. 26–64. Wiley, New York, 1986.

[12] I. Biederman, “Recognition-by-components: A theory of human image understanding,” *Psychological Review*, vol. 94, pp. 115–147, 1987.

[13] B. G. Shinn-Cunningham, “Influences of spatial cues on grouping and understanding sound,” in *Forum Acusticum 2005 Budapest*, 2005.

[14] V. Best, A. Ihlefeld, and B. Shinn-Cunningham, “The effect of auditory spatial layout in a divided attention task,” in *11th International Conference on Auditory Display (ICAD’05)*, Limerick, Ireland, 2005, pp. 17–22.

[15] V. Best, A. van Schaik, C. Jin, and S. Carlile, “Auditory spatial perception with sources overlapping in frequency and time,” *Acta Acustica United with Acustica*, vol. 91, pp. 421–428, 2005.

[16] S. H. Han and M. S. Kim, “The role of spatial working memory in visual search efficiency,” *Psychonomic Bulletin & Review*, vol. 11, pp. 275–281, 2004.

[17] G. Woodman and S. Luck, “Visual search is slowed when visuospatial working memory is occupied,” *Psychonomic Bulletin & Review*, vol. 11, pp. 269–274, 2004.

[18] G. Repovs and A. Baddeley, “The multi-component model of working memory: Explorations in experimental cognitive psychology,” *Neuroscience*, vol. 139, pp. 5–21, 2006.

[19] J. C. Webster and P. O. Thompson, “Responding both of two overlapping messages.,” *Journal of the Acoustic Society of America*, vol. 26, no. 1, pp. 396–402, 1954.

[20] J. A. Michon and J. L. Jackson, “Attentional effort and cognitive strategies in the processing of temporal information,” *Annals of the New York Academy of Sciences*, vol. 423, pp. 298–321, 1984.

[21] B. Arons, “A review of the cocktail party effect,” *Journal of the American Voice I/O Society*, vol. 12, pp. 35–50, 1992.

[22] W. Speith, J. F. Curtis, and J. C. Webster, “Responding to one of two simultaneous messages,” *Journal of the Acoustical Society of America*, vol. 26, no. 1, pp. 391–396, 1954.

[23] E. C. Cherry and W. K. Taylor, “Some further experiments on the recognition of speech, with one and two ears,” *Journal of the Acoustic Society of America*, vol. 26, pp. 554–559, 1954.

[24] E. C. Cherry, “Some experiments on the recognition of speech with one or two ears,” *Journal of the Acoustical Society of America*, vol. 25, pp. 975–979, 1953.

[25] B. Shinn-Cunningham and A. Ihlefeld, “Selective and divided attention: Extracting information from simultaneous sound sources,” in *10th International Conference on Auditory Display (ICAD’04)*, Sydney, Australia, 2004.

[26] V. Best, A. van Schaik, and S. Carlile, “Two-point discrimination in auditory displays,” in *9th International Conference on Auditory Display (ICAD’03)*, Boston, MA, USA, 2003.

[27] D. K. McGookin and S. A. Brewster, “Advantages and issues with concurrent audio presentation as part of an auditory display,” in *12th International Conference on Auditory Display (ICAD’06)*, UK, 2006.

[28] M. Puckett and D. Zicarelli, “Max/msp,” 1990-2007.

[29] J. M. Jot and O. Warusfel, “A spatial processor for musicians and sound engineers,” 1995.

[30] L. M. Brown, S. A. Brewster, R. Ramloll, M. Burton, and B. Riedel, “Design guidelines for audio presentation of graphs and tables,” in *9th International Conference on Auditory Display (ICAD’03)*, Boston, MA, USA, 2003.

[31] A. W. Mills, “On the minimum audible angle,” *Journal of the Acoustical Society of America*, vol. 30, pp. 237–246, 1958.

[32] B. C. J. Moore, *An introduction to the psychology of hearing*, Elsevier Science, USA, 2003.

[33] J. Blauert, *Spatial hearing: The psychoacoustics of human sound localization*, The MIT Press, London, England, 1997.

[34] E. Terhardt, “Pitch, consonance, and harmony,” *Journal of the Acoustical Society of America*, vol. 55, no. 5, pp. 1061–1069, 1974.

[35] B. Gardner and K. Martin, “Hrtf measurements for a kemar dummy-head microphone,” May, 1994 1994.

[36] J. G. Neuhoff, J. Wayand, and R. Knight, “Pitch change, sonification and musical expertise: Which way is up?,” in *8th International Conference on Auditory Display (ICAD’02)*, Kyoto, Japan, 2002.

[37] B. N. Walker and L. M. Mauney, “Individual differences, cognitive abilities and the interpretation of auditory graphs,” in *2004 International Conference on Auditory Display (ICAD’04)*, Sydney, Australia, 2004.

[38] H. F. Pollard and E. V. Jansson, “A tristimulus method for the specification of musical timbre,” *Acustical*, vol. 55, pp. 162–171, 1982.

[39] S. McAdams, “Recognition of sound sources and events,” in *Thinking in sound: The cognitive psychology of human audition*, S. McAdams and E. Bigand, Eds., pp. 146–198. Oxford University Press, Oxford, 1993.

[40] J. Krimphoff, S. McAdams, and S. Winsberg, “Characterisation du timbre des sons complexes. ii: Analyses acoustiques et quantification psychophysique.,” *Journal de Physique IV*, vol. 4, no. C5, pp. 625–628, 1994.

[41] P. Martignon, A. Azzali, D. Cabrera, A. Capra, and A. Farina, “Reproduction of auditorium spatial impression with binaural and stereophonic sound systems,” in *Audio Engineering Society Convention*, Barcelona, Spain, 2005.

[42] P. Zahorik, C. Tam, K. Wang, P. Bangayan, and V. Sundareswaran, “Localization accuracy in 3-d sound displays: The role of visual-feedback training,” 2003.

[43] B. G. Shinn-Cunningham, T. Streeter, and J. F. Gyss, “Perceptual plasticity in spatial auditory displays,” *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 4, pp. 418–425, 2005.

[44] B. A. Wright and M. B. Fitzgerald, “Sound-discrimination learning and auditory displays,” in *9th International Conference on Auditory Display (ICAD’03)*, Boston, MA, USA, 2003.